Asia-Pacific Journal of Science and Technology: Volume: 28. Issue: 03. Article ID.: APST-28-03-17. Research Article

APST

Asia-Pacific Journal of Science and Technology

https://www.tci-thaijo.org/index.php/APST/index

Published by the Research and Graduate Studies,
Khon Kaen University, Thailand

In silico prediction of molecular and functional annotation of hypothetical protein (ABC47680) of *Acinetobacter venetianus*

Sadia H. Tonny[1], Prodipto B. Angon[1], Sumaiya H. Omy[2,*] and Zahirul A. Talukder[2]

[1]Faculty of Agriculture, Bangladesh Agricultural University, Mymensingh, Bangladesh
[2]Plant Breeding Division, Bangladesh Agricultural Research Institute, Gazipur, Bangladesh
[*]Corresponding author: omy_pbd13@bari.gov.bd

Received 18 February 2022
Revised 22 May 2022
Accepted 18 June 2022

Abstract

Acinetobacter venetianus is a non-motile coccobacilli bacterium that is aerobic, gram-negative, has positive catalytic activity, and exhibits negative oxidative behavior. Enormous data on hypothetical or uncharacterized proteins are available in the genomic database of the bacteria. Therefore, appropriate bioinformatics tools are essential for gaining a complete understanding of the bacterial genome. Several bioinformatics approaches were used in this study to determine the prognosis of structural and functional data of the targeted hypothetical protein. An in-silico approach was used to determine several features of the hypothetical protein (HP) (ABC47680) and compare the sequences of related proteins. Various bioinformatics tools were used, including NCBI-search, MEGA-7, ExPASy ProtPram, PSORTb, SWISS-MODEL, and PSIPRED for secondary structure predictions. Physicochemical properties, subcellular localization prediction, and identification were performed, followed by subsequent examination using several approaches. An alpha helix with an extended strand was the secondary predicted structure. After validating with different servers, the tertiary structure of the protein with maximum similarity in our study was revealed. Through functional annotation, the ParE toxin superfamily was discovered as domain containing an oligonucleotide-binding (OB)-fold protein with a beta barrel, and it's a virulent protein. The OB fold protein is involved in ribonucleic acid (RNA) binding, while ParE is involved in deoxy ribonucleic acid (DNA) damage repair and environmental stress response. The protein promotes plasmid partitioning while inhibiting the process of DNA replication and cell development. Our study on in silico prediction of hypothetical protein *Acinetobacter venetianus* could be used to combat crop pests, human activities, and distinctive environmental purposes.

Keywords: *Acinetobacter venetianus*, Bioinformatics tool, Hypothetical protein, Bacterial Genome, ParE_toxin

1. Introduction

Acinetobacter venetianus was named after the Italian city of Venice, where it was discovered in the Adriatic Sea lagoon of Venice. According to its phenotypic traits, *Acinetobacter venetianus* exhibits negative oxidative behavior, strictly aerobic, gram-negative, shows positive catalytic activity, and does not have motile coccobacilli. Deoxy ribonucleic acid (DNA) guanine (G), cytosine (C) content is 34.9-47 percent, non-motile, non-fastidious, and belongs to an opportunistic pathogen with a synthetic biotechnological conveyor [1]. It acts as a source of nitrogen using ammonia and sole carbon genesis by growing in mineral media with acetate [2]. *Acinetobacter venetianus* bacteria was isolated from an oil-degrading environment, and the habitat was swallowed with activated sludge [3]. *Acinetobacter* genus was identified from several agricultural soils, forest soil, seawater [2], wetlands [4], etc. While working on finding novel characteristics of the bacteria, researchers found considerable interest in analyzing various databases of its genome. Computational biology has evolved at such a pick position that different server and web tools have been built up for predicting the function of the unidentified proteomic organism, the similarity of protein sequences, active site identification, localization in the sub-cellular organism, driving phylogenetic analysis, gene expression, motif, conserved domain analysis, and protein-protein interaction

[5]. The unknown function of any gene which is present in an organism's genome is known as hypothetical protein [6], and its existence isn't proven yet but can be visualized through an open reading frame (ORF) [7]. Around half of the proteins in the genome are classed as hypothetical proteins, which have some circumstantial evidence of genomic and proteomic features [8].

Uncharacterized protein families (UPF) have proof of existence without any characterized gene, whereas the domain of unknown functions (DUF) has no functional domain [9]. The cellular function of hypothetical protein (HP)s containing predictable structure, pathway, modeling, and sequencing could be done through proteomic analysis and homology-based annotation [10]. HPs analysis by different bioinformatics tools can define the pharmacological target for drug designing for various diseases [11]. By practicing in-silico analysis through bioinformatics tools, functional annotation of the unidentified protein can be identified.

The HPs of *Acinetobacter venetianus* (ABC47680) were used to find out biological, physicochemical, and structural functions for future up-gradation of the protein bacteria. The homology modeling technique was used to assess the model of the protein. Prediction of similarity, subcellular localization, active site detection, and protein-protein interaction was minutely analyzed. The current study's findings can be used to determine whether bacteria can survive in extreme conditions and produce toxic, biodegradable organic products. Metabolic pathway synthesis was performed along with soil remediation. *Acinetobacter* spp. elect National Health Commission (NHCs) and use them in printing dyeing industries [12]. The functional annotation can lead to a new biotechnological approach conforming to conventional annotation. A study was carried out to characterize the HPs of many bacteria species, but not on this particular bacterium [13,14]. So far, no work has been organized on the new protein that has been hypothesized through our study. This proteomic analysis will unravel the modern branch of opportunities for entomologists, and it will serve as a surfactant and an oil-degrading agent. This research is being carried out to learn more about the bacteria's functional and structural characteristics so that they can be used more successfully in human life and environmental development in the future. Because there hasn't been much detailed research on the protein of this novel developing bacteria, the purpose of this study is to conduct a systematic analysis through computational investigation on a functional protein sequence.

2. Materials and methods

2.1 Sequence retrieval and uniformity identification

Fourteen genomes are available in the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) database of *Acinetobacter venetianu*s. A HPs (ABC47680) containing 107 amino acids was selected for the study. The primary sequence of the protein was retrieved in FASTA format. A similarity search was conducted to forecast the activity of the intended suppositional protein. For finding structural similarities, NCBI database with non-redundant and SwissProt database were used through the Basic Local Alignment Search Tool (BLASTp) program [15]. A thorough procedure of performing the whole characterization process is given in the (Figure 1).

2.2 Multiple sequence alignment and phylogenetic analysis

MEGA X: Molecular Evolutionary Genetics Analysis software was used for figuring out multiple sequence alignment. The neighbor-joining approach was utilized with a 1000 replicates bootstrap test to determine the percentage of replicates [16].

2.3 Physicochemical characterization analysis

ProtParam tool of (ExPASy) was used to evaluate physicochemical characteristics [17]. The server predicted molecular weight, theoretical pI, Grand average of hydropathicity (GRAVY), aliphatic index, half-life estimation, instability index, total positively charged (TPC) aa residues (Arg + Lys), total negatively charged (TNC) aa residues (Asp + Glu).

2.4 Subcellular localization

Subcellular localization containing protein takes part in connection with cytoplasmic functioning of protein-like drug targets based on average hydrophobicity of the molecule [18]. CELLO (multi-class SVM classification system), PSORTb, SOSUIGramN, TMHMM, and HMMTOP were used to predict the hypothetical protein's subcellular location [19].

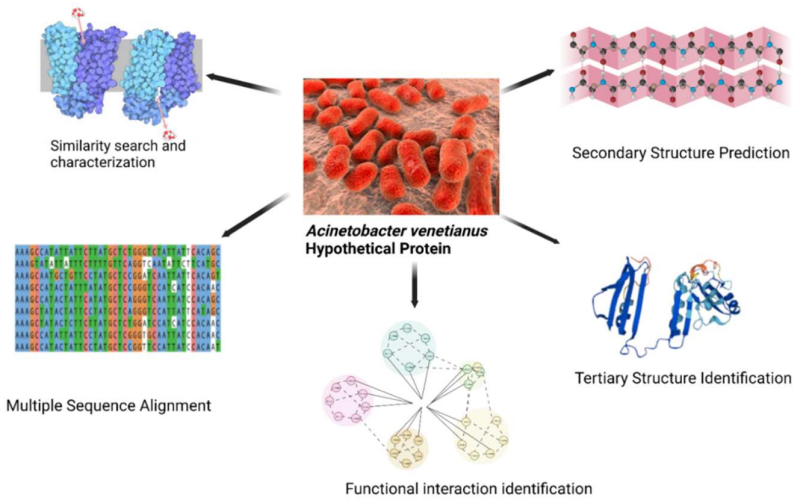

Figure 1 Summary of entire proccedure performed for annotation of the hypothetical protein.

2.5 *Conserved domain, protein-motif, evolutionary relationship and functional annotation*

For the determination of the uncharacterized protein's function, Conserved Domain Database (CDD) from NCBI [20] was checked for conserved domains. InterPro, HHpred, Pfam (https://pfam.xfam.org/) and smart server [21] were used for deciphering the function of the protein. Detailed information about protein domain, protein family for homology modelling from diverse databases were collected for analysis from Interpro and Pfam. Motif was searched at (Genome Net https://www.genome.jp/) server for protein motif search [22]. The COILS server confirmed the coiled coils within the protein. PFP-FunD SeqE Server performed for finding protein-folding pattern [23]. Along with that evolutionary history, similarities of the sequences were also found through the protein sequence and hierarchies of closely related groups of various protein families could be found also [24].

Homologous protein annotation was done manually by blast over non-redundant Uniprot sequences. Pairwise comparison from various databases by selecting template manually was done by using Hhpred server [25]. Prediction of protein family's function was performed by using CATH-Gene3D server where CATH (https://www.cathdb.info/) stands for protein structures classification and structural domain location prediction was executed by Gene3D [26].

2.6 Virulent protein sequence identification

For estimation of pathogenic potentiality, mechanisms of complex virulence of the pathogenesis can be identified through the recognition of virulent protein in bacterial protein sequence [27]. VCIMpred and VirulentPred tools were used for identifying the virulence factor of the unidentified protein sequence. VirulentPred is based on a bi-layer cascade with amino acid composition, dipeptide composition, evolutionary relationship using fivefold cross validation technique. VCIMpred is used for identifying functional protein through amino acid pattern and composition of gram-negative bacteria [28].

2.7 Prediction of secondary structure

Position-Specific Iterative Basic Local Alignment Search Tool (PSI-blast) based secondary structure prediction (PSIPRED), a 2D structure prediction server result acquired from "Position Specific Iterated-BLAST", was used to analyze the secondary structure of the hypothetical protein. The web server SOPMA (Self-Optimized Prediction Method Alignment) was also utilized to predict secondary protein structure using the homologous method based on primary sequence [29].

2.8 Tertiary structure prediction

The FASTA series of the suppositional protein was used to examine the 3D structure via the Swiss Cheese (SWISS)-MODEL server. It's based on automated comparative modeling [30]. The web server HHpred was applied based on best scoring. Robetta web tool (https://robetta.bakerlab.org/) which is automatically predicted protein structure tools [31]. Phyre2 server was also used for tertiary structure prediction using advanced remote homology detection methods.

2.9 Quality assessment

SAVES, PROCHECK, PROFUNC, and ERRAT servers were used to assess the projected tertiary structure of the hypothetical protein. Ramachandran Plot was also made by using the RAMPAGE server to find the backbone dihedral angles ψ against φ amino acid surplus in the protein composition. A server (Varify 3D) was applied to determine the compatibility of amino acids in a tertiary or 3D structure [32].

3. Results and discussion

3.1 Sequence retrieval and similarity

The physical location of the gene (locus) of the hypothetical protein (plasmid) [*Acinetobacter venetianus*] is ABC47680 with 107 amino acids. The source strain of the target protein is *Acinetobacter venetianus* VE-C3. The FASTA sequence of it is: >ABC47680.1 hypothetical protein (plasmid) [*Acinetobacter venetianus*] MYTICETPLFTKYCLVYWTQEEYEEFKTFLALNPEAGDVEPNSGGIRKIRWTSGGRGKSGGVRVIYFNR LTNGEIWLLTLYSKKQTVQLSKKTLQALVEKLNDSFND. The result of BLASTp is shown in the (Table 1 and 2).

Table 1 The output of BLASTP against UniProt/Swissprot non-redundant sequences.

Accession ID	Organism	Max Score	Total score	Query coverage (%)	Identity (%)	e-value	Accession length
WP_000288001.1	*Gammaproteobacteria*	221	221	100	100	8e-73	107
WP_004999436.1	*Acinetobacter*	220	220	100	99.07	2e-72	107
WP_109291756.1	*Acinetobacter baumannii*	219	219	100	99.07	4e-72	107
WP_074382630.1	*Acinetobacter*	219	219	100	99.07	4e-72	107
WP_152873598.1	Unclassified *Acinetobacter*	219	219	100	99.07	4e-72	107

While performing the similar search of the target protein by BLASTp, predictive query sequences similar to the protein were found. A comparison of the parameter of the maximum-total score, coverage of the query sequence, identity percentage, aa length, value was constructed. Four of the organisms are under *Acinetobacter* family and the maximum- total score varies from 219-221 with 100% query coverage which indicates that these query sequences are a good match. Identity % is around 99% on average.

Table 2 Similar protein information obtained from the UniProt database.

Entry ID	Organism	Identity (%)	Score	Value
A0A4V3D5G1	*Thiopseudomonas denitrificans*	64.4	360	1.4e-42
A0A1I3VPM5	*Paraburkholderia megapolitana*	58.7	314	1.3e-35
S6AAF4	*Sulfuricella denitrificans*	57.7	297	4.7e-33
A0A455V018	*Halomonas axialensis*	57.8	255	1e-26
Q3JED2	*Nitrosococcus oceani*	50.0	292	3.5e-32

The similar search acquired from the UniProt dataset showed a higher identity % at 64.4% with 360 scores, and the organism is *Thiopseudomonas denitrificans.* While lowest identity is 50%, with 292 score content of *Nitrosococcus oceani* organism.

3.2 Multiple sequence alignment

Using the FASTA sequences, multiple sequence alignment was performed of the unidentified protein (ABC47680.1) and the predicted homologous annotated protein. Sequence alignment of the HPs obtained from MEGA is shown in (Figure 2). A phylogenetic tree was also built by using sequence alignment and BLAST results. 0.002 distance following p-distance between branches was found with a total 107 position of 5 amino acids sequences. The complete deletion method was adopted for removing gaps and missing data.

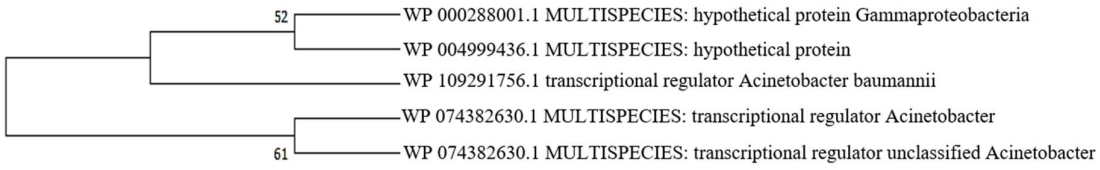

Figure 2 Phylogenetic tree with 0.002 distance using MEGA-7 Software.

3.3 Physicochemical attributes

By evaluating the amino acid properties of the hypothetical protein, physicochemical properties can be identified. This performance was conducted by the ExPASy ProtPram server. The total number of amino acids is 107 with 12338.10 Da molecular weight. Among them Leu (12), Lys (9), Thr (10), Tyr (6), Val (6), Arg (5), Asn (6), Glu (9), Gly (9), Phe (5) were in considerable number. Therefore, the highest abundant amino acid residue is 11.2% of Leu and the lowest amino acid residue is 0.9% of Met. Isoelectric point (pI) was 8.54 theoretically which demonstrates that the protein is charged negatively. The total quantity of positively charged residues (Arg + Lys) and the total quantity of negatively charged residues (Asp + Glu) are 14 and 12 respectively. 29.38 was the calculated instability index (II) which reckons the stability of the target protein. Aliphatic index and GRAVY of the unidentified protein is 81.03 and -0.427 correspondingly. This protein is polar as the value is positive. Mammalian reticulocytes half-life of the protein, in vitro condition, is 30 h, yeast (in vivo) >20 h, and E. coli (in vivo) is >10 h. The chemical formula of the protein is $C_{560}H_{868}N_{144}O_{164}S_{3,}$ and total atom number is 1739.

3.4 Subcellular localization

Determination of protein residues can be found by performing subcellular localization. Subcellular location affects functional properties, genome annotation, and determination of interaction. Enumeration of the subcellular location of the protein is in the cytoplasmic area through performing the servers CELLO 2.5, PSORTb, SOSUIGramN. Scores from PSORTb exhibited a cytoplasmic membrane score 2.00, score of outer membrane 2.00 of 2.00 score of periplasmic and extracellular cell. Eventually, the result from TMHMM 2.0 and HMMTOP computed that this protein has no transmembrane helices. This result have distinctive role in characterizing putative role in therapeutic industry.

3.5 Conserved domain, protein-motif, evolutionary relationship, and functional annotation

Primitive methods of annotation and finding the function of uncharacterized protein is to compare the similarities and dissimilarities of a standard protein sequence with the HPs data. The result of BLASTp against the nonredundant Uniprot/SwissProt database is shown in (Table 2).

The distinct function of protein in association with sequence pattern from domain and functional motif performing the biological role with folding pattern came to known to all. From CDD tool one domain was found of ParE_toxin superfamily which (accession No cI21503) consists of OB-fold protein (Figure 3). Here ParE is the toxin family of type II with anti-toxin system of which amino acid residue is 35-97 along with e-value of 6.41e-03. ParE consists of DNA sequence homology with distinct cellular mechanisms and targets [33]. Here the analysis plays a pivotal role for discovering surviving ability of the bacteria.

Pfam server analyzed the protein sequence and found 3 matches named ReIE, Phage CRI (Phage replication protein) and ATP-grasp 2. E-value of these 3 family are 2.5e-05, 0.03 and 0.16 respectively. The amino acid residue of this HPs is (19-106), (46-107) and (42-103) correspondingly. InterProScan server Toxin HigB-2 family was observed which is found in Vibrio cholerae [34]. The same output was found from the Motif server. SMART server found no hidden domain with 53-64 amino acid residue. PFP-FunDSeqE server found OB-fold (oligonucleotide/oligosaccharide-binding fold) with 5/6 stranded closed beta-barrels and 70-80 amino acid residues. In the graph, x-axis illustrate the amino acid position in the protein and the y-axis presents coiled coil along with that “window” showing the width of amino acid scanned at one time.

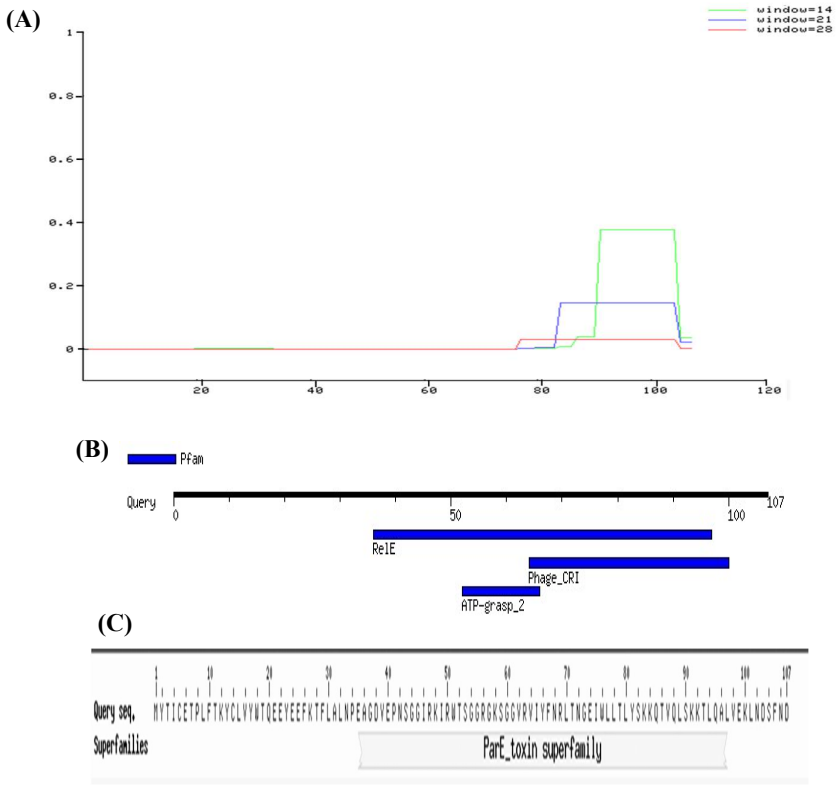

Figure 3 (A) Coils server describing amino acid residual; Windows 21(Blue), 14(Green), 28(Red); (B) Motif-Finder for functional annotation; (C) CDD for finding functional notation of the hypothetical protein.

3.6 Virulent protein sequence identification

VirulentPred tool found that the hypothetical protein is a virulent bacterium, and the score is 0.9845. Virulent bacteria can cause infection to host organisms and disease infection is occurrs. VICMpred tool analyzed a different functional group of the unidentified protein and found cellular process, metabolism, information molecule, and virulence factor 1.0145319, -0.84963815, -1.6089141, and -0.033705256 respectively. The predicted functional class is “Cellular process”.

3.7 Evaluation of secondary structure

Prediction of the secondary composition of the unidentified protein (ABC47680) was performed by SOPMA which shows the result of helix, bridge, strand, turn, coils and states of protein in association with protein interaction function [35]. The result showed that alpha helix is 35.51%, the extended strand is 26.17%, beta turn is 10.28% and the random coil is 28.04%. PRISPRED server was operated to cross-check the finding along with SYMPRED. The predicted illustrative secondary constitution of the HPs (ABC47680) is shown in (Figure 4).

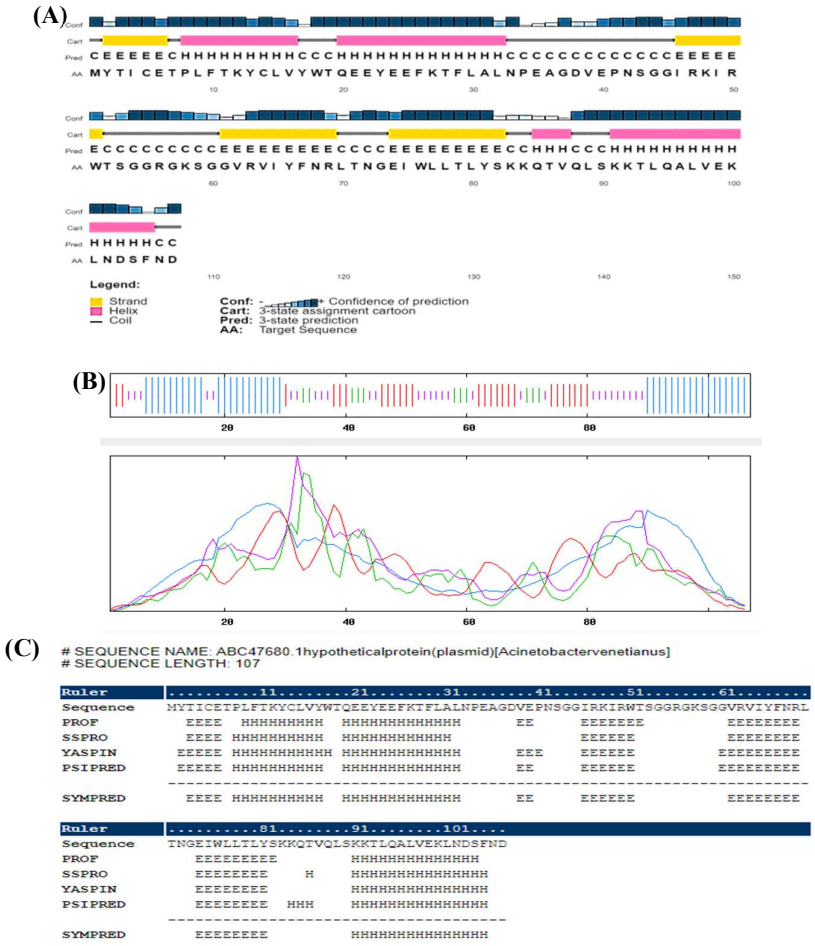

Figure 4 Predicted model of secondary structure. (A) Computation of secondary structure from PSIPRED server. (B) Structure analysis by SOPMA server. (C) Secondary structure prediction by SYMPRED Server using PSI-BLAST algorithm.

3.8 Predicted 3D structure explanation

When only amino acid sequences are available of any unidentified protein then homology modeling or comparative modeling tools are preferable to identify structural role of the protein. Meticulous information of protein-protein interaction for static conformation is determined in this analysis. More than 30% sequence identity is considered as accuracy equivalent to low-resolution X-ray structure whereas less than 30% accuracy isn't considered at all [36]. Homology modeling was done by SWISS-MODEL server with 36.89% sequence identity, 0.96 coverage and hetero-dimer oligo state (PBD ID 5ja8.2). The crystallographic resolution was 2.49 Å of the model which followed the X-ray diffraction method. Estimation of global quality, local quality, protein size residue and alignment of the model template is illustrated in (Figure 5).

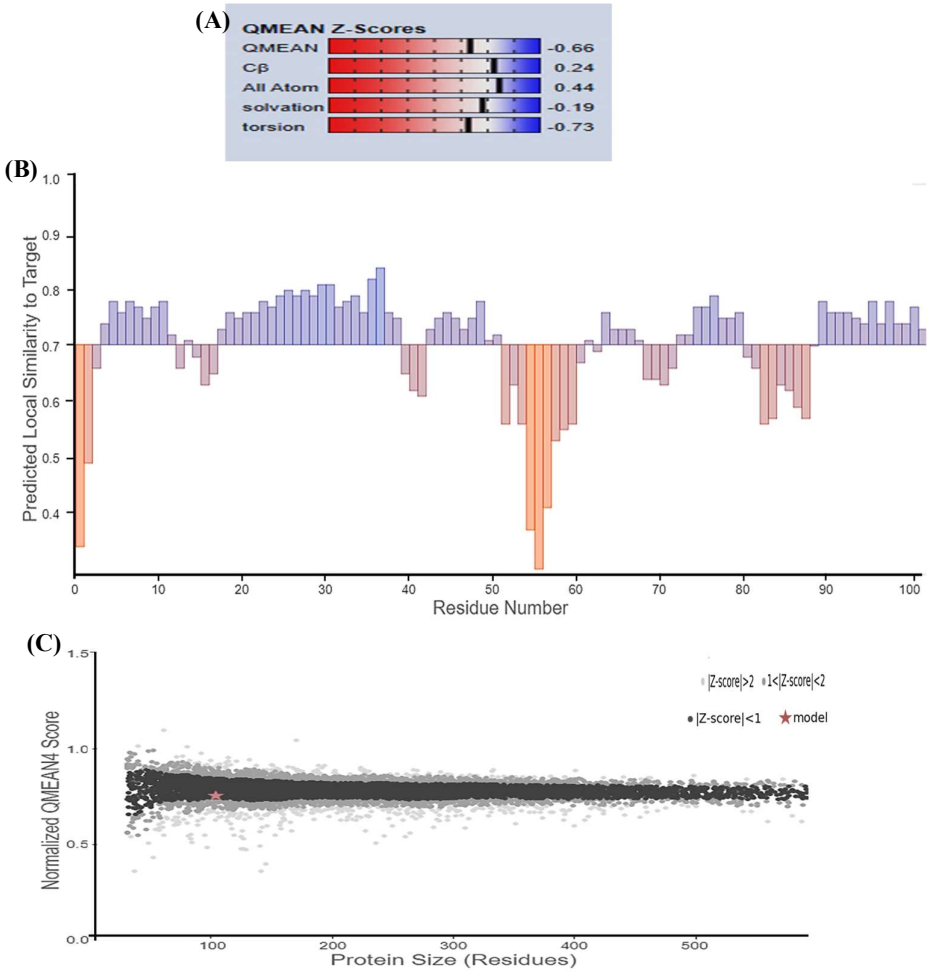

Figure 5 Evolutionary structure identification was made using BLAST for finding the tertiary structure of protein ABC47680 (A) global quality estimation, (B) local quality estimation, (C) quality comparison.

From the Robetta server confidence was found 0.83 using RoseTTAFold modeling method where 1 is good, and 0 is bad. HHpred server found 99.8% similarity of a model (PDB ID: 5MJE) with 98.59 template score. Phyre2 web server predicted 100% confidence with 88% coverage by single highest scoring template and the PDB molecule is toxin higb-2. Putative binding site for other proteins of this HPs, was predicted by building the tertiary structure of the protein. Predicted viable 3D figure of the protein is shown in (Figure 6).

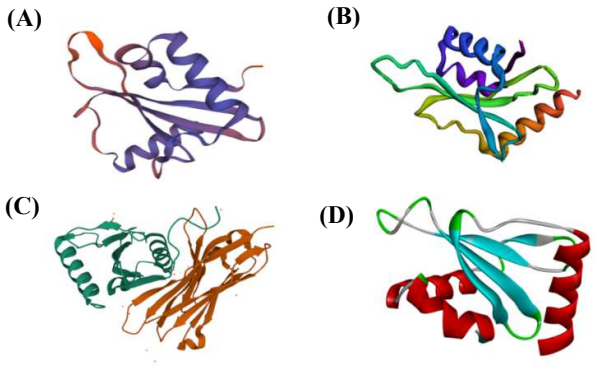

Figure 6 (A) Predicted Tertiary model using SWISS- MODEL, (B) structure from Robetta web server, (C) HHpred webtool, (D) computed structure from phyre2 server.

Robetta server predicted 3D structure with 100% accuracy which in comparison to other servers gave a more valid structure and no other more closely related server could be found. Enumerating the legitimate structure of any protein more than 90% of the residues will have to be in the favored location, score of 80% amino acid >= 0.2, 95% or higher quality factor presence [37]. In the Ramachandran plot, in the most favored region amino acid is 98.02% (Figure 7).

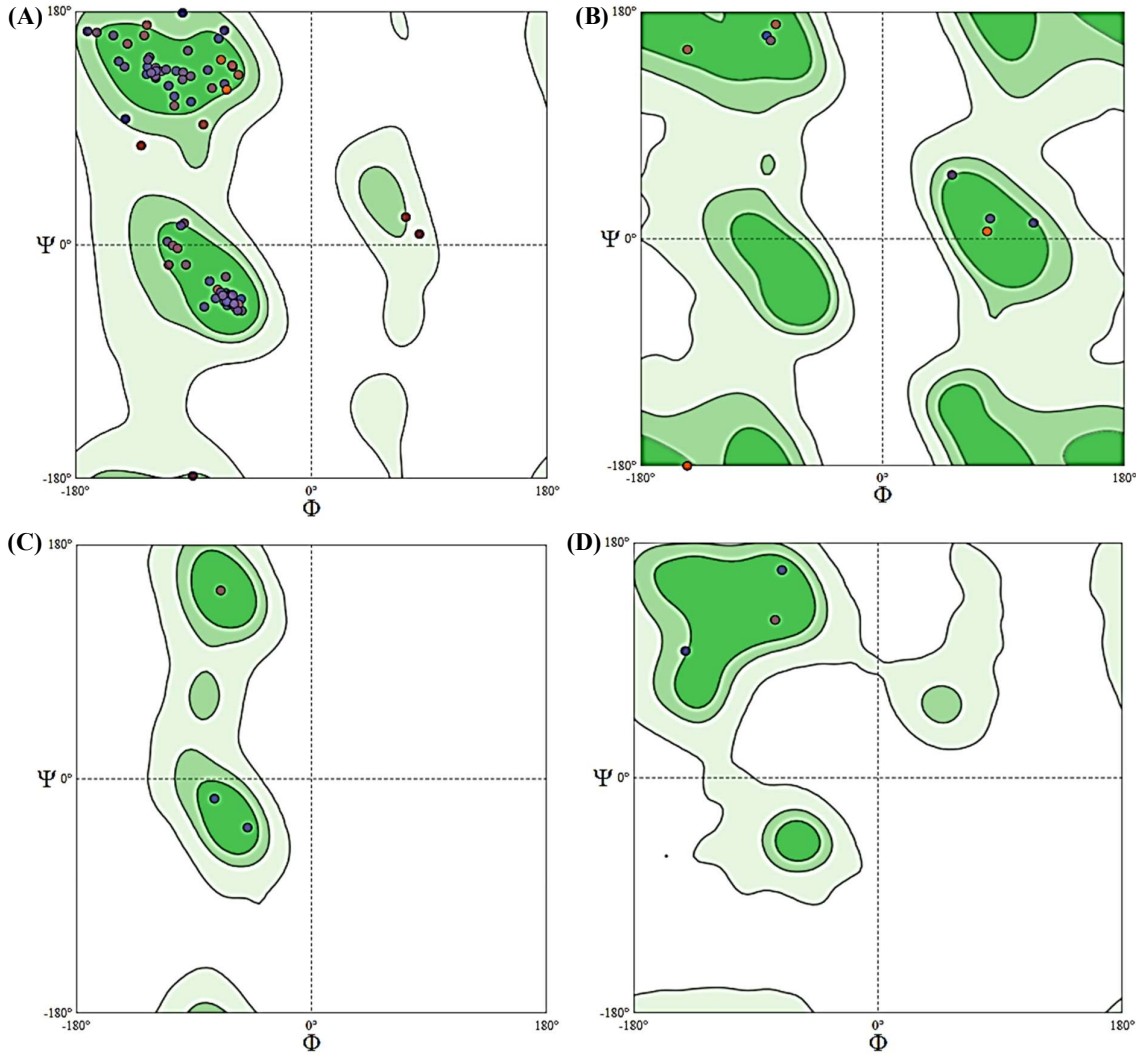

Figure 7 Ramachandran plot using SWISS server of *Acinetobacter venetianus* (A) General, (B) Glycine, (C) Proline, (D) Pre-Proline.

Based on these determinants, the model obtained from Robetta was considered an ideal model. These structures were cross-checked by using quality assessment tools from Swiss-Prot, HHpred, phyre2 and Robetta is shown in (Table 4).

Table 4 Result of quality assessment tools.

Server	ERRAT (Quality assessment)	Verify 3D (percentage of residues that have averaged 3D-1D score >= 0.2) (%)	Procheck (% residues in most favoured region) (%)
SWISS-Model	91.7973	99.93	95.6
HHpred	92.8571	91.06	87.6
Phyre2	73.3333	67.33	92
Robetta	97.8261	100	93.5

3.9 Working principle and future work prediction

ParE toxin obtained from this study, has a restricting ability of E. coli gyrases and can kill bacterial cells using quinolone antibiotic mechanism. Cell growth constriction and bacterial membrane disruption occur due to ParE toxin which is already an established study [38]. *A. venetianus* is found hazardous to the shrimp industry, hence it shouldn't grow along with shrimp production. Bacterial persistence in the type-II system is bounded by antitoxin protein which restricts the activity of the persistent protein. DNA binding proteins affect antagonistically during transcription [39].

DNA replication creates its copy which plays an important role in the growth and renewal of cells [40]. PerE toxin causes extensive cell damage by stopping DNA replication. Although it is damaging to human health, the protein can be utilized against animals that are harmful to humans because of this ability. Different types of complex chemical compounds using the protein in genomic approach insecticides can be made from this protein to eradicate harmful pests of crops. Through the application, damage of DNA inside the cells of the target species occurs and ultimately the desired target is dead. This can be a significant economical and profitable aspect of use.

4. Conclusion

The current study was carried out to identify the potential role and function of *Acinetobacter venetianus*'s undefined protein through the in-silico method in the form of a table, clarification, domain identification, phylogenetic characterization and secondary-tertiary structural view. Target uncharacterized protein revealed significant attributes like cytoplasmic mold, ParE_toxin domain and OB-fold (oligonucleotide/oligosaccharide-binding fold). The protein can stop the growth of the cell and destroy the cell with the ability to generate chemical compounds inhibiting environmental antagonists like deadly oceanic bacteria and soil inhibiting bacteria. As the bacteria thrives in humid climates and propagates very quickly through tissue culture, laboratory work can be anticipated altogether. In the future, finding the biodegrading capability, genomic activity, phenotypic nature and proper mechanism of working through molecular docking should perform for better understanding. Further use in bioremediation and bio emulsifiers can also be explored. Binding mode identification for getting interaction with plants of the bacteria should be studied.

5. Acknowledgements

Authors sincerely express their gratitude to Mst Rubaiat Nazneen Akhand, Associate Professor, Faculty of Biotechnology and Genetic Engineering, SAU for proper instruction and guideline.

6. References

[1] Fondi M, Maida I, Perrin E, Orlandini V, La Torre L, Bosi E, et al. Genomic and phenotypic characterization of the species *Acinetobacter venetianus*. Sci Rep. 2016;23(1):1-2.

[2] Vaneechoutte M, Nemec A, Musílek M, Reijden TJ, Barselaar M, Tjernberg I, et al. Description of *Acinetobacter venetianus ex* Di Cello *et al.* 1997 sp. nov. Int J Syst Evol Microbiol. 2009;59(6):1376-1381.

[3] Cello DF, Pepi M, Baldi F, Fani R. Molecular characterization of an n-alkane-degrading bacterial community and identification of a new species, *Acinetobacter venetianus*. Res Microbiol. 1997;148(3):237-249.

[4] Anandham R, Weon HY, Kim SJ, Kim YS, Kim BY, Kwon SW. *Acinetobacter brisouii* sp. nov., isolated from a wetland in Korea. J Microbiol. 2010;48(1):36-39.

[5] Gong J, Chen Y, Pu F, Sun P, He F, Zhang L, et al. Understanding membrane protein drug targets in computational perspective. Curr Drug Targets. 2019;20(5):551-564.

[6] Naveed M, Tehreem S, Usman M, Chaudhry Z, Abbas G. Structural and functional annotation of hypothetical proteins of human adenovirus: prioritizing the novel drug targets. BMC Res Notes. 2017;10(1):1-6.

[7] Bashir Z, Rizwan M, Mushtaq K, Munir A, Ali I. In silico structural and functional prediction of Phaseolus vulgaris hypothetical protein PHA VU_004G136400g. J Proteom Bioinform. 2017;10:207-216.

[8] Nimrod G, Schushan M, Steinberg DM, Ben-Tal N. Detection of functionally important regions in "hypothetical proteins" of known structure. Structure. 2008;16(12):1755-1763.

[9] Varma PB, Adimulam YB, Kodukula S. In silico functional annotation of a hypothetical protein from *Staphylococcus aureus*. J Infect Public Health. 2015;8(6):526-532.

[10] Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, Deacon AM, et al. Exploration of uncharted regions of the protein universe. PLoS Biol. 2009;7(9):e1000205.

[11] Azevedo W. Molecular dynamics simulations of protein targets identified in *Mycobacterium tuberculosis*. Curr Med Chem. 2011;18(9):1353-1366.

[12] Yang Z, Zhou J, Xu Y, Zhang Y, Luo H, Chang K, et al. Analysis of the metabolites of indole degraded by an isolated *Acinetobacter pittii* L1. Biomed Res Int. 2017;2017:2564363.

[13] Ferdous N, Reza MN, Emon MT, Islam MS, Mohiuddin AK, Hossain MU. Molecular characterization and functional annotation of a hypothetical protein (SCO0618) of *Streptomyces coelicolor* A3 (2). Genomics Inform. 2020;18(3):e28.

[14] Rahman A, Susmi TF, Yasmin F, Karim M, Hossain MU. Functional annotation of an ecologically important protein from *Chloroflexus aurantiacus* involved in polyhydroxyalkanoates (PHA) biosynthetic pathway. SN Appl Sci. 2020;2(11):1-3.

[15] Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36 Suppl 2:W5-W9.

[16] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547.

[17] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The proteomics protocols handbook. Berlin: Springer Science & Business Media; 2005. p. 571-607.

[18] Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, et al. In silico subtractive genomics for target identification in human bacterial pathogens. Drug Dev Res. 2011;72(2):162-177.

[19] Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006;64(3):643-651.

[20] Bauer MA, Anderson JB, Derbyshire MK, Scott DWC, Gonzales NR, Gwadz M, et al. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res. 2007;35 Suppl 1:D237-D240.

[21] Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol. 2001;313(4):903-919.

[22] Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res. 2002;30(1):42-46.

[23] Shen HB, Chou KC. Predicting protein fold pattern with functional domain and sequential evolution information. J Theor Biol. 2009;256(3):441-446.

[24] Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2015;43(D1):D213-D221.

[25] Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33 Suppl 1:W244-W248.

[26] Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, Lee D, et al. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res. 2012;41(D1):D490-D498.

[27] Garg A, Gupta D. VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinform. 2008;9(1):1-2.

[28] Saha S, Raghava GP. VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition. Genom Proteom Bioinform. 2006;4(1):42-47.

[29] Lin K, Simossis VA, Taylor WR, Heringa J. A simple and fast secondary structure prediction method using hidden neural networks. Bioinformatics. 2005;21(2):152-159.

[30] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 2003;31(13):3381-3385.

[31] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32 Suppl 1:W526-W531.

[32] Kihara D, Chen H, Yang YD. Quality assessment of protein structure models. Curr Protein Pept Sci. 2009;10(3):216-228.

[33] Kamruzzaman M, Iredell J. A ParDE-family toxin antitoxin system in major resistance plasmids of *Enterobacteriaceae* confers antibiotic and heat tolerance. Sci Rep. 2019;9(1):1-2.

[34] Budde PP, Davis BM, Yuan J, Waldor MK. Characterization of a higBA toxin-antitoxin locus in *Vibrio cholerae*. J Bacteriol. 2007;189(2):491-500.

[35] Canduri F, Peres PC, Caceres RA, Azevedo J, Filgueira W. CDK9 a potential target for drug development. Med Chem. 2008;4(3):210-218.

[36] Xiang Z. Advances in homology protein structure modeling. Curr Protein Pept Sci. 2006;7(3):217-227.

[37] Hooda V, babu Gundala P, Chinthala P. Sequence analysis and homology modeling of peroxidase from Medicago sativa. Bioinformatics. 2012;8(20):974.

[38] Jiang Y, Pogliano J, Helinski DR, Konieczny I. ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of *Escherichia coli* gyrase. Mol Microbiol. 2002;44(4):971-979.

[39] Mutschler H, Meinhart A. ε/ζ systems: their role in resistance, virulence, and their potential for antibiotic development. J Mol Med. 2011;89(12):1183-1194.

[40] Singh J. Role of DNA replication in establishment and propagation of epigenetic states of chromatin. Semin Cell Dev Biol. 2014;30:131-143.